A P P L I C A T I O N

for

UNITED STATES LETTERS PATENT

on

2,5-DKG PERMEASES

by

Véronique A. Dartois
James A. Hoch
Fernando Valle
and
Manoj Kumar

Sheets of Drawings: four (4)

Docket No.: P-SR 4877

CERTIFICATE OF MAILING BY "EXPRESS MAIL"

"EXPRESS MAIL" MAILING LABEL NUMBER: EL 856 980 504 US

DATE OF DEPOSIT: August 3, 2001

I HEREBY CERTIFY THAT THIS PAPER OR FEE IS BEING DEPOSITED WITH THE UNITED STATES POSTAL SERVICE "EXPRESS MAIL POST OFFICE TO ADDRESSEE" SERVICE UNDER 37 C.F.R. 1.10 ON THE DATE INDICATED ABOVE, AND IS ADDRESSED TO: COMMISSIONER FOR PATENTS WASHINGTON, D.C. 20231.

Patrick Lee
(TYPED OR PRINTED NAME OR PERSON MAILING PAPER OR FEE)

Patrick Lee
(SIGNATURE OF PERSON MAILING PAPER OR FEE)

Attorneys
CAMPBELL & FLORES LLP
4370 La Jolla Village Drive, Suite 700
San Diego, California 92122
USPTO CUSTOMER NO. 23601

2,5-DKG PERMEASES

This application claims the benefit of U.S. Provisional Application No. 60/___(yet to be assigned), filed August 4, 2000, which was converted from U.S. Serial No. 09/633,294, and U.S. Provisional Application No. 60/___ (yet to be assigned), filed September 29, 2000, which was converted from U.S. Serial Application No. 09/677,032, each of which is incorporated herein by reference in its entirety.

This invention was made in part with U.S. Government support under Cooperative Agreement 70NANB5H1138 and ATP NIST project Identification Number 1995-05-0007E. The U.S. Government has certain rights in this invention.

BACKGROUND OF THE INVENTION

FIELD OF THE INVENTION

The present invention relates generally to microbial transporter proteins and, more specifically, to novel 2,5-diketo-D-gluconic acid (2,5-DKG) permeases.

BACKGROUND INFORMATION

Adequate intake of ascorbic acid, or vitamin C, is recognized as an important factor in maintaining health. To ensure adequate intake of ascorbic acid, the chemical is now added to many foods, drinks and cosmetic products, and is also sold as a direct vitamin supplement. To meet the commercial demand for ascorbic acid, there is a need to develop more efficient processes for its production.

Although there are a number of alternative methods of producing ascorbic acid, one of the least expensive and most ecologically sound methods is biofermentation. Bacterial strains have now been engineered to express all of the enzymes required for the stepwise conversion of an inexpensive sugar source, such as D-glucose, to a stable precursor of ascorbic acid, 2-keto-L-gulonic acid (2-KLG) (see U.S. Patent No. 5,032,514 and references therein). 2-KLG can be readily converted to ascorbic acid by chemical or enzymatic procedures.

Figure 2 shows schematically the enzymatic reactions that take place in the bioconversion of D-glucose to 2-KLG. As shown in Figure 2, the enzymatic reactions that lead from D-glucose, to D-gluconic acid, to 2-keto-D-gluconic acid (2-KDG), to 2,5-diketo-D-gluconic acid (2,5-DKG), take place at the surface of the bacterial cell. 2,5-DKG must then enter the cell in order for its enzymatic conversion to 2-KLG.

Much effort has been expended in increasing the efficiency of the enzymatic reactions involved in 2-KLG production. For example, U.S. Patent No. 5,032,514 describes methods for increasing 2-KLG production by reducing metabolic diversion of 2,5-DKG to products other than 2-KLG.

Increasing uptake of 2,5-DKG by a bacterial strain suitable for biofermentation could be advantageous in increasing 2-KLG production. Expressing additional copies of an endogenous 2,5-DKG permease, or expressing an exogenous 2,5-DKG permease with superior properties, could

increase uptake of 2,5-DKG. However, to date, no 2,5-DKG permease has been identified or characterized that could be used in this manner.

Therefore, there exists a need to identify and characterize nucleic acid molecules encoding 2,5-DKG permeases, so that permeases with advantageous properties can be used in the commercial production of ascorbic acid and in other important applications. The present invention satisfies this need and provides related advantages as well.

SUMMARY OF THE INVENTION

The invention provides an isolated nucleic acid molecule encoding a polypeptide which has 2,5-DKG permease activity. In one embodiment, the isolated nucleic acid molecule contains a nucleotide sequence having at least 40% identity to a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9 and 11. In another embodiment, the isolated nucleic acid molecule contains a nucleotide sequence which encodes a polypeptide having at least 40% identity to an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10 and 12.

Also provided are vectors and cells containing isolated nucleic acid molecules encoding polypeptides having 2,5-KDG permease activity. In one embodiment, the cells are bacterial cells selected from the genera *Pantoea* and *Klebsiella*.

The invention also provides methods of identifying and isolating nucleic acid molecules encoding polypeptides

which have 2,5-DKG permease activity. Also provided are methods of enhancing 2-KLG production, by expressing the nucleic acid molecules of the invention in suitable bacterial cells.

Further provided are isolated polypeptides having 2,5-DKG permease activity, and immunogenic peptides therefrom. The invention also provides antibodies specific for such polypeptides and peptides.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A and 1B shows an alignment of the amino acid sequences of the 2,5-DKG permeases designated YiaX2 (SEQ ID NO:12); PE1 (SEQ ID NO:2); PE6 (SEQ ID NO:4); prmA (SEQ ID NO:8); prmB (SEQ ID NO:10) and PK1 (SEQ ID NO:6).

Figure 2 shows the biosynthetic pathway from glucose to 2-KLG in a bacterial strain suitable for biofermentation.

Figure 3 shows the metabolic selection strategy used to identify novel 2,5-DKG permeases.

DETAILED DESCRIPTION OF THE INVENTION

The invention provides novel nucleic acid molecules encoding polypeptides having 2,5-DKG permease activity, and related products and methods. The molecules of the invention can advantageously be used to increase the efficiency of 2-KLG bioproduction, and thus to lower the cost of commercial ascorbic acid production.

Naturally occurring 2,5-DKG permeases are polypeptides localized to the cytoplasmic membrane of microorganisms, which are predicted, using commercially available topology prediction programs, to contain about 10 to 12 transmembrane domains. Each transmembrane spanning segment is about 20 amino acids in length, with the intracellular and extracellular loops ranging from about 2 to about 83 amino acids in length. Generally, the loop between the fifth and sixth transmembrane domain spanning segments is larger than the other loops.

Naturally occurring 2,5-DKG permeases are typically about 350-550 amino acids in length, such as about 400-450 amino acids in length, and particularly about 425-440 amino acids in length.

The nucleotide sequences encoding six exemplary 2,5-DKG permeases are set forth as follows, with the designation and organismal source of the molecule indicated in parentheses: SEQ ID NO:1 (PE1 from an environmental source); SEQ ID NO:3 (PE6 from an environmental source); SEQ ID NO:5 (PK1 from *Klebsiella oxytoca*); SEQ ID NO:7 (prmA from *Pantoea citrea*); SEQ ID NO:9 (prmB from *Pantoea citrea*); and SEQ ID NO:11 (YiaX2 from *Klebsiella oxytoca*). The corresponding encoded 2,5-DKG permease amino acid sequences are set forth as SEQ ID NO:2 (PE1); SEQ ID NO:4 (PE6); SEQ ID NO:6 (PK1); SEQ ID NO:8 (prmA); SEQ ID NO:10 (prmB); and SEQ ID NO:12 (YiaX2).

2,5-DKG permeases from different microorganisms exhibit extensive amino acid sequence relatedness over their entire length, as is evidenced by the six-way sequence alignment shown in Figure 1. The overall identity of the six permeases shown in Figure 1 is about 17%, and the overall similarity, taking into account conservative substitutions, is about 43%.

Based on their predicted topological and sequence similarity, 2,5-DKG permeases disclosed herein can be further subdivided into two structural families. The three 2,5-DKG permeases designated YiaX2, PE6 and PrmA are representative of one family of permeases, sharing about 50% overall identity in a three-way amino acid sequence alignment. The three 2,5-DKG permeases designated PK1, PE1 and PrmB are representative of a second family of related permeases, sharing about 60% overall identity in a three-way amino acid sequence alignment.

Naturally occurring 2,5-DKG permeases also exhibit 2,5-DKG permease activity. The term "2,5-DKG permease activity," as used herein, refers to the ability of the polypeptide, when expressed in its native orientation at the cell membrane, to transport 2,5-DKG across the cytoplasmic membrane, in comparison with an unrelated control polypeptide. Such transport can be either unidirectional or bidirectional.

2,5-DKG permease activity can be determined by a variety of methods. For example, 2,5-DKG permease activity can be determined using a metabolic selection assay, as described further in the Example, below. Briefly, a

bacterial cell either naturally deficient in 2,5-DKG permease activity, or made deficient in 2,5-DKG permease activity, is identified or produced. As described in the Example, bacterial cells can be made deficient in endogenous 2,5-DKG permease activity by preparing a deletion mutant of one or more endogenous 2,5-DKG permease genes, using the polymerase chain reaction, following methods known in the art. The term "deficient," as used in relation to a cell deficient in 2,5-DKG permease activity, is intended to refer to endogenous 2,5-DKG permease activity that is comparable to, or less than, the endogenous permease activity of a *K. oxytoca* strain deleted in the *yiaX2* gene, such as the strain *K. oxytoca* Δ*yiaX2* [*tkr idnO*], as assessed either by a growth assay or by a 2,5-DKG uptake assay.

A cell useful in a metabolic selection assay to determine 2,5-DKG permease activity of an expressed polypeptide can further naturally be capable of converting intracellular 2,5-DKG to carbon and energy, or made capable of such conversion by recombinant expression of appropriate metabolic enzymes. As described in the Example, a combination of nucleic acid molecules encoding a 2-keto-reductase (*tkr*) and a 5-keto-reductase (*idnO*), from any bacterial species, can be expressed in the cell, which together provide the cell with the ability to catalyze the reduction of 2,5-DKG to gluconic acid. Gluconic acid can then be used by the cell as a carbon and energy source that supports cell growth.

An exemplary bacterial cell suitable for metabolic assays to determine 2,5-DKG permease activity is the strain *K. oxytoca* Δ*yiaX2* [*tkr idnO*] shown in Figure 3 and

described in the Example, below. This strain has a deleted *yiaX2* 2,5-DKG permease gene, and also recombinantly expresses the *tkr/idnD/idnO* operon set forth as SEQ ID NO:13 on a high copy number plasmid. Within SEQ ID NO:13, nucleotides 292-1236 encode a 2-keto-reductase (tkr) (SEQ ID NO:14); nucleotides 1252-2280 encode an idonic acid dehydrogenase (idnD) (SEQ ID NO:15); and nucleotides 2293-3045 encode a 5-keto-reductase (idnO) (SEQ ID NO:16). Alternatively, nucleic acid molecules encoding polypeptides which contain modifications from the amino acid sequences designated SEQ ID NO:14 or 16, but which retain 2-keto-reductase activity or 5-keto-reductase activity, respectively, can be used in metabolic assays. Exemplary amino acid sequences have at least 60%, such as at least 70%, preferably 80%, 90%, 95% or greater identity to SEQ ID NOS:14 or 16, respectively.

The ability of such a bacterial cell to grow on medium containing 2,5-DKG as the sole carbon source, upon expression of a candidate 2,5-DKG permease, is a measure of the ability of the expressed permease to transport 2,5-DKG into the cell, and is thus a measure of its 2,5-DKG permease activity. Each of SEQ ID NOS:2, 4, 6, 8, 10 and 12 was demonstrated to have 2,5-DKG permease activity, as evidenced by the ability of *K. oxytoca* Δ*yiaX2* [*tkr idnO*] expressing each permease to grow on 2,5-DKG as the sole carbon source.

Likewise, 2,5-DKG permease activity can be determined by measuring uptake of labeled or unlabeled 2,5-DKG. For example, 2,5-DKG can be detectably labeled, such as with a fluorescent or radioactive tag. The ability of a cell or membrane vesicle expressing a 2,5-DKG permease to

take up the detectable label when provided with detectably labeled 2,5-DKG, can be determined using detection assays specific for the particular label, which are well known in the art. Likewise, uptake of unlabeled 2,5-DKG can be measured by HPLC or other sensitive detection assay known in the art. Uptake of 2,5-DKG is thus a measure of permease activity. Each of SEQ ID NOS:2, 4, 6, 8, 10 and 12 exhibits 2,5-DKG permease activity as determined by assay of uptake of radiolabeled 2,5-DKG by bacterial cells expressing the recombinant permeases.

Additionally, 2,5-DKG permease activity can be measured in any cell in which 2,5-DKG can be converted to a product, by measuring production of the product in the presence of extracellular 2,5-DKG. For example, in a cell naturally expressing, or recombinantly expressing, a 2,5-DKG reductase, intracellular 2,5-DKG is converted to 2-KLG. The ability of the bacterial cell to produce 2-KLG when provided with extracellular 2,5-DKG, upon expression of a 2,5-DKG permease, is a measure of the ability of the expressed permease to transport 2,5-DKG into the cell, and is thus a measure of its 2,5-DKG permease activity. Intracellular 2-KLG can be detected, for example, using HPLC or other sensitive detection methods known in the art. Other metabolic products of 2,5-DKG can also be detected, by similar methods.

It will be appreciated that a variety of alternative assays can be used to determine 2,5-DKG permease activity. For instance, the change in pH across a cell or vesicle membrane as 2,5-DKG, an acid, is transported across

the membrane can be detected. Similarly, a decrease over time in extracellular 2,5-DKG can be determined.

Accordingly, using any of the activity assays described herein, those skilled in the art can distinguish between a polypeptide having 2,5-DKG permease activity, and a polypeptide not having such activity.

A 2,5-DKG permease of the invention can selectively transport 2,5-DKG. As used herein in relation to transport activity, the term "selective" refers to preferential transport of 2,5-DKG rather than 2-KLG into or out of the cell. A permease that selectively transports 2,5-DKG will transport 2,5-DKG at least 2-fold, such as at least 5-fold, including greater than 10-fold more efficiently than it transports 2-KLG. A permease that selectively transports 2,5-DKG is particularly advantageous in applications where it is desirable to increase intracellular production of 2-KLG, such as in the commercial production of ascorbic acid. In particular, employing a permease that selectively transports 2,5-DKG prevents intracellular 2-KLG from competing with extracellular 2,5-DKG for permease-mediated transport through the membrane, and increases the overall efficiency of intracellular 2-KLG production.

It will be appreciated that the assays described above for determining 2,5-DKG permease activity can be modified to simultaneously, or separately, determine 2-KLG permease activity. For example, a metabolic assay can be designed in which a bacterial cell can convert either intracellular 2,5-DKG or 2-KLG to carbon and energy. In

such a cell, the relative ability of the cell to grow on 2,5-DKG as the sole carbon source, compared with its ability to grow on 2-KLG as the sole carbon source, is a measure of the ability of the expressed permease to selectively transport 2,5-DKG. Using such an assay, it was determined that the 2,5-DKG permeases designated YiaX2, PE1, PE6, prmA and prmB are non-selective for 2,5-DKG, as they also efficiently catalyze the transport of 2-KLG, as *K. oxytoca* Δ*yiaX2* [*tkr idnO*] cells expressing such permeases grow well on either 2,5-DKG or 2-KLG. In contrast, PK1 selectively transports 2,5-DKG, and *K. oxytoca* Δ*yiaX2* [*tkr idnO*] cells expressing PK1 (SEQ ID NO:6) grow on 2,5-DKG but not on 2-KLG as the sole carbon source.

The invention provides an isolated nucleic acid molecule encoding a polypeptide which has 2,5-DKG permease activity. The invention nucleic acid molecules of the invention are suitable for a variety of commercial and research applications. For example, one or more of the invention nucleic acid molecules can be expressed in bacterial cells in order to enhance the rate of uptake of 2,5-DKG by the cells. Enhancing uptake of 2,5-DKG has a variety of applications, such as in commercial production of 2,5-DKG itself, or commercial production of any metabolic product of 2,5-DKG. For example, 2,5-DKG uptake is a rate limiting step in the biosynthesis of 2-KLG, which is a stable intermediate in the synthesis of ascorbic acid. 2-KLG can thus be obtained from bacterial cells expressing 2,5-DKG permeases, and converted to ascorbic acid.

Additionally, the invention nucleic acid molecules can be used as probes or primers to identify and isolate

2,5-DKG permease homologs from additional species, or as templates for the production of mutant permeases, using methods known in the art and described further below. Such permeases can have advantageous properties compared with the 2,5-DKG permeases disclosed herein as SEQ ID NOS:2, 4, 6, 8, 10 and 12, such as greater enzymatic activity or greater 2,5-DKG selectivity.

In one embodiment, an isolated nucleic acid molecule of the invention is not completely contained within the nucleotide sequence designated SEQ ID NO:19 of WO 00/22170, which is the *K. oxytoca yia* operon. In another embodiment, the isolated nucleic acid molecule of the invention is not completely contained within the nucleotide sequence herein designated SEQ ID NO:11. In another embodiment, the encoded polypeptide is not completely contained within the amino acid sequence herein designated SEQ ID NO:12.

The term "isolated," as used herein, is intended to mean that the molecule is altered, by the hand of man, from how it is found in its natural environment. For example, an isolated nucleic acid molecule can be a molecule operatively linked to an exogenous nucleic acid sequence. An isolated nucleic acid molecule can also be a molecule removed from some or all of its normal flanking nucleic acid sequences, such as removed from one or more other genes within the operon in which the nucleic acid molecule is normally found.

Specifically with respect to an isolated nucleic acid molecule containing the nucleotide sequence designated SEQ ID NO:11, or encoding the *yiaX2* polypeptide designated SEQ ID NO:12, the term "isolated" is intended to mean that the nucleic acid molecule does not contain any of the flanking open reading frames (orfs) present in the *K. oxytoca yia* operon, such as the orfs designated *lyxK* and *orf1*, described in WO 00/22170.

An isolated molecule can alternatively, or additionally, be a "substantially pure" molecule, in that the molecule is at least 60%, 70%, 80%, 90 or 95% free from cellular components with which it is naturally associated. An isolated nucleic acid molecule can be in any form, such as in a buffered solution, a suspension, a heterologous cell, a lyophilized powder, or attached to a solid support.

The term "nucleic acid molecule" as used herein refers to a polynucleotide of natural or synthetic origin. A nucleic acid molecule can be single- or double-stranded genomic DNA, cDNA or RNA, and represent either the sense or antisense strand or both. A nucleic acid molecule can thus correspond to the recited sequence, to its complement, or both.

The term "nucleic acid molecule" is intended to include nucleic acid molecules that contain one or more non-natural nucleotides, such as nucleotides having modifications to the base, the sugar, or the phosphate portion, or having one or more non-natural linkages, such as phosphothioate linkages. Such modifications can be

advantageous in increasing the stability of the nucleic acid molecule, particularly when used in hybridization applications.

Furthermore, the term "nucleic acid molecule" is intended to include nucleic acid molecules modified to contain a detectable moiety, such as a radiolabel, a fluorochrome, a ferromagnetic substance, a luminescent tag or a detectable binding agent such as biotin. Nucleic acid molecules containing such moieties are useful as probes for detecting the presence or expression of a 2,5-DKG permease nucleic acid molecule.

In one embodiment, the isolated nucleic acid molecule encoding a polypeptide which has 2,5-DKG permease activity contains a nucleotide sequence comprising nucleotides 1-20, 1-100, 101-120, 101-200, 201-220, 201-300, 301-320, 301-400, 401-420, 401-500, 501-520, 501-600, 601-620, 601-700, 701-720, 701-800, 801-820, 801-900, 901-920, 901-1000, 1001-1020, 1001-1100, 1101-1120, 1100-1200, 1201-1220, 1201-1300, 1301-1320, 1301-1400, 1401-1420 or 1401-1500 of any of SEQ ID NOS:1, 3, 5, 7, 9 or 11.

In another embodiment, the isolated nucleic acid molecule encoding a polypeptide which has 2,5-DKG permease activity encodes an amino acid sequence comprising amino acids 1-10, 1-50, 51-60, 51-100, 101-110, 101-150, 151-160, 151-200, 201-210, 201-250, 251-260, 251-300, 301-310, 301-350, 351-361, 351-400, 401-410 or 401-439 of any of SEQ ID NOS:2, 4, 6, 8, 10 or 12.

In one embodiment, the isolated nucleic acid molecule encoding a polypeptide which has 2,5-DKG permease activity contains a nucleotide sequence having at least 35% identity to any of the 2,5-DKG permease nucleic acid molecules designated SEQ ID NOS:1, 3, 5, 7, 9 or 11. Preferably, such a molecule will have at least 40% identity to any of these recited SEQ ID NOS, such as at least 45%, 50%, 60%, 70% or 80% identity, including at least 90%, 95%, 98%, 99% or greater identity to SEQ ID NOS:1, 3, 5, 7, 9 or 11.

In another embodiment, the isolated nucleic acid molecule encoding a polypeptide which has 2,5-DKG permease activity contains a nucleotide sequence which encodes a polypeptide having at least 35% identity to any of the 2,5-DKG permease polypeptides designated SEQ ID NOS:2, 4, 6, 8, 10 or 12. Preferably, the encoded polypeptide will have at least 40% identity to any of these recited SEQ ID NOS, such as at least 45%, 50%, 60%, 70%, 80% identity, including at least 90%, 95%, 98%, 99% or greater identity to SEQ ID NOS:2, 4, 6, 8, 10 or 12.

The term "percent identity" with respect to a nucleic acid molecule or polypeptide of the invention is intended to refer to the number of identical nucleotide or amino acid residues between the aligned portions of two sequences, expressed as a percent of the total number of aligned residues, as determined by comparing the entire sequences using a CLUSTAL V computer alignment and default parameters. CLUSTAL V alignments are described in Higgens, Methods Mol. Biol. 25:307-318 (1994), and an exemplary

CLUSTAL V alignment of 2,5-DKG permease amino acid sequences is presented in Figure 1.

Due to the degeneracy of the genetic code, the nucleotide sequence of a native nucleic acid molecule can be modified and still encode an identical or substantially similar polypeptide. Thus, degenerate variants of SEQ ID NOS:1, 3, 5, 7, 9 or 11 are exemplary invention nucleic acid molecules encoding polypeptides having 2,5-DKG permease activity.

Additionally, nucleic acid molecules encoding 2,5-DKG permeases from other species of microorganisms are exemplary invention nucleic acid molecules. The six permeases designated YiaX2, PE1, PE6, prmA, prmB and PK1, which were isolated from at least three, and likely four, different species of microorganisms, share substantial nucleotide sequence identity. For example, the two most similar of the disclosed 2,5-DKG permease nucleotide sequences, SEQ ID NO:5 (PK1) and SEQ ID NO:1 (PE1), share 86% identity across their length. The two most dissimilar of the disclosed 2,5-DKG permease nucleotide sequences, SEQ ID NO:7 (prmA) and SEQ ID NO:9 (prmB), share 51% identity across their length. In contrast, a search of GenBank reveals no other nucleotide sequences, including sequences which encode transporter proteins and other transmembrane proteins, that exhibit significant identity or similarity to any of the disclosed 2,5-DKG permease nucleotide sequences over the entire length of their sequences.

The six permeases disclosed herein also share substantial amino acid sequence identity over their entire

length, as described previously. For example, PK1 from *Klebsiella oxytoca* (SEQ ID NO:6), and PE1, from an environmental source (SEQ ID NO:2), are 93% identical at the amino acid level. The amino acid sequence in the GenBank database most closely related to a 2,5-DKG permease, which is a putative tartrate transporter from *Agrobacterium vitis* (GenBank Accession U32375 or U25634) is 33% identical to SEQ ID NO:12 (YiaX2), and shares less identity with the other disclosed 2,5-DKG permeases. Other sequences with some degree of identity in the GenBank database to the disclosed 2,5-DKG permease include membrane transporter proteins from a variety of species, including phthalate transporter proteins from *B. cepacia* (AF152094) and *P. putida* (D13229); hydroxyphenylacetate transporters from *S. dublin* (AF144422) and *E. coli* (Z37980); and probable transporter proteins from *S. coelicolor* (AL136503 and AL132991) each of which has about 27% or less identity at the amino acid level to the recited SEQ ID NOS.

In view of the high degree of identity between different 2,5-DKG permease nucleic acid molecules and encoded polypeptides within a single species and between different microbial species, additional 2,5-DKG permeases from other species can be readily identified and tested. Thus, nucleic acid molecules of the invention include nucleic acid molecules that encode polypeptides having 2,5-DKG permease activity from any microbial species. Microorganisms that contain 2,5-DKG permeases can be recognized by their ability to actively transport 2,5-DKG, such that they can grow on 2,5-DKG as the sole carbon source, or incorporate 2,5-DKG in an uptake assay. Such microorganisms can include, for example, bacteria, including

Archaebacteria, gram positive and gram negative bacteria; yeast; and fungi.

Exemplary bacteria which contain 2,5-DKG permeases include Proteobacteria, and more specifically Enterobacteria and Pseudomonads (e.g. *P. aeruginosa*), as described in the Example. Exemplary Enterobacteria include species from the genera *Klebsiella* (e.g. *K. oxytoca*, from which SEQ ID NOS:5 and 11 were obtained) and *Pantoea* (e.g. *P. citrea*, from which SEQ ID NOS:7 and 9 were obtained, and *P. agglomerans*). Sources of such microorganisms include public repositories, such as the American Type Culture Collection (ATCC), and commercial sources. It will be appreciated that the taxonomy and nomenclature of bacterial genera are such that the same or similar strains are sometimes reported in the literature as having different names. For example, *Klebsiella oxytoca* (e.g. ATCC 13182) has alternatively been described as *Aerobacter aerogenes, Klebsiella aerogenes* and *Klebsiella pneumoniae*. Likewise, *Pantoea agglomerans* (e.g. ATCC 21998) has alternatively been described as *Erwinia herbicola* and *Acetomonas albosesamae*. The terms "*Klebsiella*" and "*Pantoea*," as used herein, are intended to refer to the genera of the strains deposited as ATCC 13182 and 21998, respectively.

Additionally, microorganisms from which 2,5-DKG permease nucleic acid molecules can be obtained are microorganisms present in environmental samples. For example, the 2,5-DKG permease nucleic acid molecules designated SEQ ID NOS:1 and 3 were obtained from environmental samples. As used herein, the term "environmental sample" refers to a sample obtained from

natural or man-made environments, which generally contains a mixture of microorganisms.

Exemplary environmental samples are samples of soil, sand, freshwater or freshwater sediments, marine water or marine water sediments, industrial effluents, hot springs, thermal vents, and the like. Within an environmental sample there are likely to be microorganisms that are unidentified, and also microorganisms that are uncultivable. Isolation of invention 2,5-DKG permease molecules of the invention from microorganisms present in environmental samples does not require either identification or culturing of the microorganism.

Furthermore, nucleic acid molecules of the invention include nucleic acid molecules encoding amino acid sequences that are modified by one or more amino acid additions, deletions or substitutions with respect to the native sequence of SEQ ID NOS:2, 4, 6, 8, 10 or 12. Such modifications can be advantageous, for example, in enhancing the stability, expression level, enzymatic activity, or 2,5-DKG selectivity of the permease. If desired, such modifications can be randomly generated, such as by chemical mutagenesis, or directed, such as by site-directed mutagenesis of a native permease sequence, using methods well known in the art.

An amino acid sequence that is modified from a native permease amino acid sequence can include one or more conservative amino acid substitutions, such as substitution of an apolar amino acid with another apolar amino acid (such as replacement of leucine with an isoleucine, valine,

alanine, proline, tryptophan, phenylalanine or methionine); substitution of a charged amino acid with a similarly charged amino acid (such as replacement of a glutamic acid with an aspartic acid, or replacement of an arginine with a lysine or histidine); or substitution of an uncharged polar amino acid with another uncharged polar amino acid (such as replacement of a serine with a glycine, threonine, tyrosine, cysteine, asparagine or glutamine). A modified amino acid sequence can also include one or more nonconservative substitutions without adversely affecting the desired biological activity.

Computer programs known in the art can provide guidance in determining which amino acid residues can be substituted without abolishing the enzymatic activity of a 2,5-DKG permease (see, for example, Eroshkin et al., Comput. Appl. Biosci. 9:491-497 (1993)).

Additionally, guidance in modifying amino acid sequences while retaining or enhancing functional activity is provided by aligning homologous 2,5-DKG permease polypeptides from various species (see Figure 1). It is well known in the art that evolutionarily conserved amino acid residues and domains are more likely to be important for maintaining biological activity than less well-conserved residues and domains. Thus, it would be expected that substituting a residue which is highly conserved among the six 2,5-DKG permeases shown in Figure 1 (or among the members of the two structural families of permeases, defined as SEQ ID NOS:2, 6 and 10, and SEQ ID NOS:4, 8 and 12) with a non-conserved residue may be deleterious, whereas making the same substitution at a residue which varies widely among

the different permeases would likely not have a significant effect on biological activity.

A comparison of the amino acid sequences of PE1 (SEQ ID NO:2), which transports both 2,5-DKG and 2-KLG, and PK1 (SEQ ID NO:6), which selectively transports 2,5-DKG, indicates that the regions responsible for 2,5-DKG selectivity must reside in the 7% of amino acids which differ between these two sequences. Therefore, modifying all or some of these differing residues in a 2,5-DKG permease to those found in the PK1 sequence would be expected to increase 2,5-DKG selectivity of the permease.

Alignment of the six 2,5-DKG permeases described herein also provides guidance as to regions where additions and deletions are likely to be tolerated. For example the N and C termini, and the region around amino acids 225-250 (based on the numbering of SEQ ID NO:12 (yiaX2)) appear to be regions that are relatively tolerant of amino acid insertions and deletions, as evidenced by gaps in the sequence alignment. Modified 2,5-DKG permeases can thus include "tag" sequences at such sites, such as epitope tags, histidine tags, glutathione-S-transferase (GST) and the like, or sorting sequences. Such additional sequences can be used, for example, to facilitate purification or characterization of a recombinant 2,5-DKG permease.

It will be appreciated that confirmation that any particular nucleic acid molecule is a nucleic acid molecule of the invention can be obtained by determining the 2,5-DKG permease activity of the encoded polypeptide, using one or more of the functional assays described herein.

The invention further provides an isolated nucleic acid molecule encoding a polypeptide which has 2,5-DKG permease activity, wherein the nucleic acid molecule is operatively linked to a promoter of gene expression. The term "operatively linked," as used herein, is intended to mean that the nucleic acid molecule is positioned with respect to either the endogenous promoter, or a heterologous promoter, in such a manner that the promoter will direct the transcription of RNA using the nucleic acid molecule as a template.

Methods for operatively linking a nucleic acid to a desired promoter are well known in the art and include, for example, cloning the nucleic acid into a vector containing the desired promoter, or appending the promoter to a nucleic acid sequence using PCR. A nucleic acid molecule operatively linked to a promoter of RNA transcription can be used to express 2,5-DKG transcripts and polypeptides in a desired host cell or *in vitro* transcription-translation system.

The choice of promoter to operatively link to an invention nucleic acid molecule will depend on the intended application, and can be determined by those skilled in the art. For example, if a particular gene product may be detrimental to a particular host cell, it may be desirable to link the invention nucleic acid molecule to a regulated promoter, such that gene expression can be turned on or off. An exemplary inducible promoter known in the art is the *lacPO* promoter/operator, which is repressed by the *lacI*^{q} gene product provided by certain host cells, and induced in the presence of 0.01 to 1 mM IPTG (see Example, below). For

other applications, weak or strong constitutive promoters may be preferred.

The invention further provides a vector containing an isolated nucleic acid molecule encoding a polypeptide which has 2,5-DKG permease activity. The vectors of the invention will generally contain elements such as a bacterial origin of replication, one or more selectable markers, and one or more multiple cloning sites. The choice of particular elements to include in a vector will depend on factors such as the intended host cell or cells; whether expression of the inserted sequence is desired; the desired copy number of the vector; the desired selection system, and the like. The factors involved in ensuring compatibility between a host and a vector for different applications are well known in the art.

In applications in which the vectors will be used for recombinant expression of the encoded polypeptide, the isolated nucleic acid molecules will generally be operatively linked to a promoter of gene expression, as described above, which may be present in the vector or in the inserted nucleic acid molecule. In cloning and subcloning applications, however, promoter elements need not be present.

An exemplary vector suitable for both cloning applications and for expressing 2,5-DKG permeases in different bacterial species is the low copy number plasmid pCL1920 described by Lerner et al., Nucleic Acids Res. 18:4621 (1994), which contains a spectinomycin resistance gene (see Example, below).

Also provided are cells containing an isolated nucleic acid molecule encoding a polypeptide which has 2,5-DKG permease activity. The isolated nucleic acid molecule will generally be contained within a vector compatible with replication in the particular host cell. However, for certain applications, incorporation of the nucleic acid molecule into the bacterial genome will be preferable.

The cells of the invention can be any cells in which a 2,5-DKG permease will be expressed and folded into an active conformation. Guidance in choosing appropriate host cells is provided by identifying cell types which express other functional 10 to 12 transmembrane transporter proteins. For example, 10 to 12 transmembrane transporter proteins are found in a variety of bacterial species, as well as in yeast (e.g. *S. pombe*), *Arabidopsis*, and *Drosophila*. Therefore, depending on the particular application for the host cell, a host cell of the invention can be a bacterial cell, yeast, Arabidopsis or Drosophila cell.

In a preferred embodiment, the cell is a bacterial cell. The choice of bacterial cell will depend on the intended application. For example, for routine subcloning applications, the cell can be any convenient laboratory strain of bacteria, such as *E. coli*, which can be transformed with the isolated nucleic acid molecules and vectors of the invention by methods well known in the art.

For assessment of encoded 2,5-DKG permease activity, the cell can be a bacterial strain suitable for

metabolic assays, such as a strain which endogenously expresses, or which is engineered to express, enzymes that catalyze the conversion of 2,5-DKG to essential products. An exemplary strain suitable for metabolic assays is the *K. oxytoca* Δ*yiaX2* [*tkr idnO*] strain designated MGK002[pDF33] described further in the Example, below, which provides for the conversion of intracellular 2,5-DKG to gluconic acid, which can be used as a carbon and energy source.

For use in the commercial bioproduction of 2,5-DKG metabolites, the cell can be a bacterial strain which endogenously expresses, or which is engineered to express, a 2,5-DKG reductase. As described in U.S. Patent No. 5,032,514, 2,5-DKG reductases are found in genera including Brevibacterium, Arthrobacter, Micrococcus, Staphylococcus, Pseudomonas, Bacillus, Citrobacter and Corynebacterium. Therefore, a cell of the invention can be a bacterial cell of any of these genera, or a bacterial cell engineered to express a 2,5-DKG reductase of any of these genera.

A cell able to produce 2,5-DKG metabolites will preferably also be able to catalyze the extracellular production of 2,5-DKG from an inexpensive carbon source, such as glucose. An exemplary pathway from D-glucose to 2,5-DKG involves the enzymatic conversion of D-glucose to D-gluconic acid (catalyzed by D-glucose dehydrogenase), from D-gluconic acid to 2-Keto-D-gluconic acid (catalyzed by D-gluconate dehydrogenase), and from 2-Keto-D-gluconic acid to 2,5-DKG (catalyzed by 2-Keto-D-gluconic acid dehydrogenase), as is shown in Figure 2. These steps can be carried out by organisms of several genera, including Gluconobacter, Acetobacter and Erwinia (also called Pantoea).

A bacterial cell useful for the production of 2-KLG from D-glucose is the *Pantoea aggolmerans* (also referred to as *Erwinia herbicola* or *Acetomonas albosesamae*) strain described in U.S. Patent No. 5,032,514, designated ATCC 21998 ptrp 1-35 tkrAΔ3, or a derivative of this strain with improved properties. Contemplated improvements to this strain, which can be produced by genetic engineering, include deletion of enzymes that divert glucose to metabolites other than 2-KLG, such that yield of 2-KLG is increased. Other contemplated improvements to this strain include mutations that provide for improved recovery and purification of 2-KLG.

The *Pantoea* strain described in U.S. Patent No. 5,032,514 recombinantly expresses a 2,5-DKG reductase from Corynebacterium (described in U.S. Patent No. 4,757,012). The strain further contains a mutation that results in a non-functional *tkrA* gene and is thus deficient in 2-keto reductase activity. Mutation of the tkrA gene is advantageous in reducing metabolic diversion of 2-KLG to L-idonic acid, and metabolic diversion of 2,5-DKG to 5-keto-D-gluconate from 2-KLG.

Expression of one or more 2,5-DKG permeases of the invention in such cells significantly increases overall production of 2-KLG from D-glucose, which lowers the cost of commercial production of ascorbic acid.

The cells of the invention can contain one, two or more isolated nucleic acid molecules of the invention that encode polypeptides having 2,5-DKG permease activity. For example, the cell can contain an isolated nucleic acid

molecule encoding at least one polypeptide having at least 80% identity to any of SEQ ID NOS:2, 4, 6, 8, 10 or 12, and optionally will contain two or more such nucleic acid molecules, in any combination. Preferably, at least one such encoded polypeptide selectively transports 2,5-DKG.

In a preferred embodiment, a bacterial cell of the invention suitable for bioproduction of 2-KLG contains an isolated nucleic acid molecule encoding a polypeptide having at least 95% identity to the 2,5-DKG selective permease designated SEQ ID NO:8 (prmA); and optionally further containing at least one isolated nucleic acid molecule encoding a polypeptide having at least 95% identity to a 2,5-DKG permease selected from the group consisting of SEQ ID NO:4 (PE6), SEQ ID NO:10 (prmB) and SEQ ID NO:6 (PK1).

The invention also provides a method of enhancing production of 2-KLG. The method consists of culturing a bacterial cell, wherein the cell contains an isolated nucleic acid molecule encoding a polypeptide which has 2,5-DKG permease activity, under conditions wherein the encoded 2,5-DKG permease is expressed and intracellular 2,5-DKG is converted to 2-KLG. Cells suitable for this purpose, such as the Pantoea strain described in U.S. Patent No. 5,032,514, have been described above. Optionally, the 2-KLG so produced can be chemically or enzymatically converted to a desired product such as ascorbic acid, following methods known in the art.

The invention further provides isolated oligonucleotide molecules that contain at least 17 contiguous nucleotides from any of the nucleotide sequences

referenced as SEQ ID NOS:1, 3, 5, 7, 9 or 11. As used herein, the term "oligonucleotide" refers to a nucleic acid molecule that contains at least 17 contiguous nucleotides from the reference sequence and which may, but need not, encode a functional protein. Thus, an oligonucleotide of the invention can contain at least 18, 19, 20, 22 or 25 contiguous nucleotides, such as at least 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 250, 300, 400, 500, 750, 1000 or more contiguous nucleotides from the reference nucleotide sequence, up to the full length of the reference nucleotide sequence. The oligonucleotides of the invention are thus of sufficient length to be useful as sequencing primers, PCR primers, hybridization probes or antisense reagents, and can also encode polypeptides having 2,5-DKG permease activity, or immunogenic peptides therefrom. Those skilled in the art can determine the appropriate length and sequence of an oligonucleotide of the invention for a particular application.

For certain applications, such as for detecting 2,5-DKG expression in a cell or library, it will be desirable to use isolated oligonucleotide molecules of the invention that specifically hybridize to a nucleic acid molecule encoding a 2,5-DKG permease. The term "specifically hybridize" refers to the ability of a nucleic acid molecule to hybridize, under stringent hybridization conditions as described below, to a nucleic acid molecule that encodes a 2,5-DKG permease, without hybridizing to a substantial extent under the same conditions with nucleic acid molecules that do not encode 2,5-DKG permeases, such as unrelated molecules that fortuitously contain short regions of identity with a permease sequence. Thus, a nucleic acid

molecule that "specifically hybridizes" is of a sufficient length and contains sufficient distinguishing sequence from a 2,5-DKG permease to be characteristic of the 2,5-DKG permease. Such a molecule will generally hybridize, under stringent conditions, as a single band on a Northern blot or Southern blot prepared from mRNA of a single species.

As used herein, the term "stringent conditions" refers to conditions equivalent to hybridization of a filter-bound nucleic acid molecule to a nucleic acid in a solution containing 50% formamide, 5X Denhart's solution, 5X SSPE, 0.2% SDS at 42°C, followed by washing the filter in 0.1X SSPE, and 0.1% SDS at 65°C twice for 30 minutes. Equivalent conditions to the stringent conditions set forth above are well known in the art, and are described, for example in Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, New York (1992).

Nucleotide sequences that are characteristic of each of SEQ ID NOS:1, 3, 5, 7 or 9, or which are common to two, three or more of SEQ ID NOS:1, 3, 5, 7, 9 or 11 can readily be determined by aligning the sequences using a CLUSTAL V alignment program. Oligonucleotides containing regions which are common to two or more different 2,5-DKG permease nucleic acid molecules can advantageously be used as PCR primers or hybridization probes to isolate or detect nucleic acid molecules encoding 2,5-DKG permeases from other species.

The oligonucleotides of the invention can, but need not, encode polypeptides having 2,5-DKG activity. Thus, the invention oligonucleotides can contain sequences

from the 5' or 3' untranslated region, or both, of the nucleotide sequences designated SEQ ID NOS:1, 3, 5, 7, 9 or 11, or contain coding sequences, or both. As described above with respect to the term "nucleic acid molecule," the invention oligonucleotides can be derived from either the sense or antisense strand of the recited SEQ ID NO.

The oligonucleotides of the invention can also advantageously be used to direct the incorporation of amino acid additions, deletions or substitutions into a recombinant 2,5-DKG permease. In such applications, it will be understood that the invention oligonucleotides can contain nucleotide modifications with respect to SEQ ID NOS:1, 3, 5, 7, 9 or 11 such that the oligonucleotides encode the desired amino acid modifications to SEQ ID NOS:2, 4, 6, 8, 10 or 12, so long as they contain at least 17 contiguous residues from the reference sequence.

Exemplary oligonucleotides of the invention are oligonucleotides that contain a sequence selected from nucleotides 1-20, 1-100, 101-120, 101-200, 201-220, 201-300, 301-320, 301-400, 401-420, 401-500, 501-520, 501-600, 601-620, 601-700, 701-720, 701-800, 801-820, 801-900, 901-920, 901-1000, 1001-1020, 1001-1100, 1101-1120, 1100-1200, 1201-1220, 1201-1300, 1301-1320, 1301-1400, 1401-1420 or 1401-1500 of any of SEQ ID NOS:1, 3, 5, 7, 9 or 11.

The invention further provides a kit containing a pair of 2,5-DKG permease oligonucleotides packaged together, either in a single container or separate containers. The pair of oligonucleotides are preferably suitable for use in PCR applications for detecting or amplifying a nucleic acid

molecule encoding a 2,5-DKG permease. The kit can further contain written instructions for use of the primer pair in PCR applications, or solutions and buffers suitable for such applications.

The invention further provides isolated oligonucleotides that contain a nucleotide sequence encoding a peptide having at least 10 contiguous amino acids of an amino acid selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10 or 12. Such oligonucleotides can encode at least 10, 12, 15, 20, 25 or more contiguous amino acids of SEQ ID NOS:2, 4, 6, 8, 10 or 12, such as at least 30, 40, 50, 75, 100, 200, 300, 400 or more contiguous amino acids from the reference sequence. The encoded peptides can be expressed from such oligonucleotides, by routine methods, and used to produce, purify or characterize 2,5-DKG antibodies, as will be discussed further below. The peptides encoded by such oligonucleotides can, but need not, additionally have 2,5-DKG permease enzymatic activity.

In one embodiment, the isolated oligonucleotide encodes an amino acid sequence selected from amino acids 1-10, 1-50, 51-60, 51-100, 101-110, 101-150, 151-160, 151-200, 201-210, 201-250, 251-260, 251-300, 301-310, 301-350, 351-361, 351-400, 401-410, 401-439 of any of SEQ ID NOS:2, 4, 6, 8, 10 or 12.

Isolated nucleic acid molecules which encode polypeptides having 2,5-DKG permease activity, as well as the isolated oligonucleotides described above, will be subsequently referred as "2,5-DKG permease nucleic acid molecules."

The isolated 2,5-DKG permease nucleic acid molecules of the invention can be prepared by methods known in the art. The method chosen will depend on factors such as the type and size of nucleic acid molecule one intends to isolate; whether or not it encodes a biologically active polypeptide (e.g. a polypeptide having permease activity or immunogenicity); and the source of the nucleic acid molecule. Those skilled in the art can isolate or prepare 2,5-DKG permease nucleic acid molecules as genomic DNA or desired fragments therefrom; as full-length cDNA or desired fragments therefrom; or as full-length mRNA or desired fragments therefrom, from any microorganism of interest.

An exemplary method of preparing a 2,5-DKG permease nucleic acid molecule is by isolating a recombinant construct which encodes and expresses a polypeptide having 2,5-DKG permease activity. As described in the Example, one useful method is to provide a metabolic selection system where bacterial cell growth is made dependent on expression of a 2,5-DKG permease, introducing expressible DNA, such as a cDNA or genomic library, into the assay cells, selecting surviving cells under the selective conditions, and isolating the introduced DNA. Alternatively, a screening method can be designed, such that a cell will exhibit a detectable signal only when expressing a functional 2,5-DKG permease. An exemplary detectable signal is intracellular incorporation of a detectable label present on 2,5-DKG. Additionaly screening and selection strategies suitable for identifying nucleic acid molecules encoding metabolic enzymes are described, for example, in PCT publication WO 00/22170 and U.S. Patent Nos. 5,958,672 and 5,783,431.

A further method for producing an isolated 2,5-DKG permease nucleic acid molecule involves amplification of the nucleic acid molecule using 2,5-DKG permease-specific primers and the polymerase chain reaction (PCR). Using PCR, a 2,5-DKG permease nucleic acid molecule having any desired boundaries can be amplified exponentially starting from as little as a single gene or mRNA copy, from any cell having a 2,5-DKG permease gene.

Given the high degree of identity among the six disclosed 2,5-DKG permeases, those skilled in the art can design suitable primers for isolating additional 2,5-DKG permease nucleic acid molecules. Such primers are preferably degenerate oligonucleotides that encode, or are complementary to, short consensus amino acid sequences present in two or more of the 2,5-DKG permeases disclosed herein, such as oligonucleotides that encode 10 or more contiguous amino acids present in at least two of SEQ ID NOS: 2, 6 and 10, or oligonucleotides that encode 10 or more contiguous amino acids present in at least two of SEQ ID NOS:4, 8 and 12. Such sequences can be determined from an alignment of amino acid sequences shown in Figure 1. Exemplary amino acid sequences present in at least two of SEQ ID NOS:2, 6 and 10 are amino acids 19-31, 115-124, 146-156, and 339-348 of SEQ ID NO:2. Exemplary amino acid sequences present in at least two of SEQ ID NOS:4, 8 and 12 are amino acids 55-64, 60-69, 252-261, and 370-379 of SEQ ID NO:8.

Methods are well known in the art to determine or modify PCR reaction conditions when using degenerate primers to isolate a desired nucleic acid molecule. The amplified

product can subsequently be sequenced, used as a hybridization probe, or used for 5' or 3' RACE to isolate flanking sequences, following procedures well known in the art and described, for example, in Ausubel et al., Current Protocols in Molecular Biology, John Wiley & Sons, New York (2000).

Given the high degree of sequence identity and structural relatedness among the six disclosed 2,5-DKG permeases, homologs from any other species can readily be identified by either hybridization or antibody screening. For example, an isolated 2,5-DKG permease nucleic acid molecule can be identified by screening a library, such as a genomic library, cDNA library or expression library, with a detectable nucleic acid molecule or antibody. Such libraries are commercially available from a variety of microorganisms, or can be produced from any available microorganism or environmental sample of interest using methods described, for example, in PCT publication WO 00/22170. The library clones identified as containing 2,5-DKG permease nucleic acid molecules can be isolated, subcloned and sequenced by routine methods.

Furthermore, 2,5-DKG permease nucleic acid molecules can be produced by direct synthetic methods. For example, a single stranded nucleic acid molecule can be chemically synthesized in one piece, or in several pieces, by automated synthesis methods known in the art. The complementary strand can likewise be synthesized in one or more pieces, and a double-stranded molecule made by annealing the complementary strands. Direct synthesis is particularly advantageous for producing relatively short

molecules, such as oligonucleotide probes and primers, and also for producing nucleic acid molecules containing modified nucleotides or linkages.

The invention also provides an isolated polypeptide which has 2,5-DKG permease activity. Such isolated polypeptides, when expressed in their normal configuration at the cell membrane, are useful in applications in which enhanced uptake of 2,5-DKG is desirable, such as in bioproduction of 2-KLG. The isolated polypeptides of the invention can also be added to a culture medium, preferably in a membrane vesicle, to compete with membrane-bound permeases for 2,5-DKG, and thus to stop 2,5-DKG uptake. Thus, isolated polypeptides having 2,5-DKG permease activity can be used to regulate production of 2,5-DKG metabolites.

In one embodiment, an isolated polypeptide of the invention is not encoded by a nucleotide sequence completely contained within the nucleotide sequence designated SEQ ID NO:19 of WO 00/22170, which is the *K. oxytoca yia* operon. In another embodiment, an isolated polypeptide of the invention is not completely contained within the amino acid sequence herein designated SEQ ID NO:12.

An "isolated" polypeptide of the invention is altered by the hand of man from how it is found in its natural environment. For example, an isolated 2,5-DKG permease can be a molecule that is recombinantly expressed, such that it is present at a higher level in its native host, or is present in a different host. Alternatively, an "isolated" 2,5-DKG permease of the invention can be a

substantially purified molecule. Substantially purified 2,5-DKG permeases can be prepared by methods known in the art. Specifically with respect to a polypeptide encoding the *yiaX2* polypeptide designated SEQ ID NO:12, the term "isolated" is intended to mean that polypeptide is not present in association with the polypeptides expressed by other genes in the *K. oxytoca yia* operon, such as the genes designated *lyxK* and *orf1,* described in WO 00/22170.

In one embodiment, an isolated polypeptide having 2,5-DKG permease activity contains an amino acid sequence having at least 40% identity to an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10 or 12. Preferably, the encoded polypeptide will have at least 45% identity to any of the recited SEQ ID NOS, such as at least 50%, 60%, 70%, 80% identity, including at least 90%, 95%, 98%, 99% or greater identity.

In another embodiment, the isolated polypeptide having 2,5-DKG permease activity contains at least 10 contiguous amino acids of any of SEQ ID NOS:2, 4, 6, 8, 10 or 12. Exemplary invention polypeptides contain an amino acid sequence of amino acids 1-10, 1-50, 51-60, 51-100, 101-110, 101-150, 151-160, 151-200, 201-210, 201-250, 251-260, 251-300, 301-310, 301-350, 351-361, 351-400, 401-410, 401-439 of any of SEQ ID NOS:2, 4, 6, 8, 10 or 12.

Also provided is an isolated immunogenic peptide having an amino acid sequence derived from a 2,5-DKG permease. Such isolated immunogenic peptides are useful, for example, in preparing and purifying 2,5-DKG antibodies. The term "immunogenic," as used herein, refers to a peptide

that either is capable of inducing 2,5-DKG permease-specific antibodies, or capable of competing with 2,5-DKG permease-specific antibodies for binding to a 2,5-DKG permease. Peptides that are likely to be immunogenic can be predicted using methods and algorithms known in the art and described, for example, by Irnaten et al., Protein Eng. 11:949-955 (1998), and Savoie et al., Pac. Symp. Biocomput. 1999:182-189 (1999). The immunogenicity of the peptides of the invention can be confirmed by methods known in the art, such as by delayed-type hypersensitivity response assays in an animal sensitized to a 2,5-DKG permease, or by direct or competitive ELISA assays.

An isolated immunogenic peptide of the invention can contain at least 10 contiguous amino acids of a polypeptide selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10 or 12, such as amino acids 1-10, 1-50, 51-60, 51-100, 101-110, 101-150, 151-160, 151-200, 201-210, 201-250, 251-260, 251-300, 301-310, 301-350, 351-361, 351-400, 401-410, 401-439 of any of SEQ ID NOS:2, 4, 6, 8, 10 or 12. Such a peptide can have at least 12, 15, 20, 25 or more contiguous amino acids of the reference sequence, including at least 30, 40, 50, 75, 100, 200, 300, 400 or more contiguous amino acids from the reference sequence, up to the full-length sequence.

For the production of antibodies that recognize 2,5-DKG permeases in their native configuration, such peptides will preferably contain at least part of an extracellular or intracellular domain from the permease. An extracellular or intracellular domain is generally characterized by containing at least one polar or positively

or negatively charged residue, whereas a transmembrane domain is generally characterized as an uninterrupted stretch of about 20 contiguous hydrophobic residues. Commercially available computer topology programs can be used to determine whether a peptide is likely to correspond to an extracellular or intracellular domain or to a transmembrane region. Immunogenic peptides of the invention derived from a transmembrane region are useful to raise antibodies for use in applications such as immunoblotting, where the 2,5-DKG polypeptide need not be in its native configuration to be recognized.

The structural and functional characteristics and applications of 2,5-DKG permease polypeptides of the invention have been described above with respect to the encoding nucleic acid molecules, and are equally applicable in reference to the isolated polypeptides of the invention. Isolated polypeptides having 2,5-DKG permease activity, as well as the isolated immunogenic peptides of the invention, will subsequently be referred to as "2,5-DKG permeases."

Methods for recombinantly producing 2,5-DKG permeases have been described above with respect to nucleic acid molecules, vectors and cells of the invention. 2,5-DKG permeases can alternatively be prepared by biochemical procedures, by isolating membranes from bacteria that naturally express, or recombinantly express, 2,5-DKG permeases. The membranes can be further fractionated by size or affinity chromatography, electrophoresis, or immunoaffinity procedures, to achieve the desired degree of purity. Purification can be monitored by a variety of

procedures, such as by immunoreactivity with 2,5-DKG permease antibodies, or by a functional assay.

Immunogenic peptides can be produced from purified or partially purified 2,5-DKG permease polypeptides, for example, by enzymatic or chemical cleavage of the full-length polypeptide. Methods for enzymatic and chemical cleavage and for purification of the resultant peptide fragments are well known in the art (see, for example, Deutscher, Methods in Enzymology, Vol. 182, "Guide to Protein Purification," San Diego: Academic Press, Inc. (1990)).

Alternatively, 2,5-DKG permeases can be produced by chemical synthesis. If desired, such as to optimize their functional activity, stability or bioavailability, such chemically synthesized molecules can include D-stereoisomers, non-naturally occurring amino acids, and amino acid analogs and mimetics. Sawyer, Peptide Based Drug Design, ACS, Washington (1995) and Gross and Meienhofer, The Peptides: Analysis, Synthesis, Biology, Academic Press, Inc., New York (1983). For certain applications, such as for detecting the polypeptide, it can also be useful to incorporate one or more detectably labeled amino acids into a chemically synthesized permease, such as radiolabeled or fluorescently labeled amino acids.

An isolated 2,5-DKG permease of the invention can further be conjugated to carrier molecules, such as keyhole lympet hemocyanin, which can enhance recognition by the immune system of the isolated 2,5-DKG permease for production of antibodies. For certain applications, such as

to increase the stability or bioactivity of the molecule, or to facilitate its identification, the 2,5-DKG permease can be chemically or enzymatically derivatized, such as by acylation, phosphorylation or glycosylation.

The invention also provides an antibody specific for a polypeptide having 2,5-DKG permease activity, such as an antibody specific for a polypeptide having the amino acid sequence of any of SEQ ID NOS:2, 4, 6, 8, 10 or 12. Also provided is an antibody specific for an isolated peptide that contains at least 10 contiguous amino acids of any of SEQ ID NOS:2, 4, 6, 8, 10 or 12, wherein the peptide is immunogenic. The antibodies of the invention can be used, for example, to detect or isolate 2,5-DKG permeases from expression libraries or cells.

The term "antibody," as used herein, is intended to include molecules having specific binding activity for a 2,5-DKG permease of at least about 1×10^5 M^{-1}, preferably at least 1×10^7 M^{-1}, more preferably at least 1×10^9 M^{-1}. The term "antibody" includes both polyclonal and monoclonal antibodies, as well as antigen binding fragments of such antibodies (e.g. Fab, $F(ab')_2$, Fd and Fv fragments and the like). In addition, the term "antibody" is intended to encompass non-naturally occurring antibodies, including, for example, single chain antibodies, chimeric antibodies, bifunctional antibodies, CDR-grafted antibodies and humanized antibodies, as well as antigen-binding fragments thereof.

Methods of preparing and isolating antibodies, including polyclonal and monoclonal antibodies, using

peptide and polypeptide immunogens, are well known to those skilled in the art and are described, for example, in Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1988). Non-naturally occurring antibodies can be constructed using solid phase peptide synthesis, can be produced recombinantly or can be obtained, for example, by screening combinatorial libraries consisting of variable heavy chains and variable light chains. Such methods are described, for example, in Huse et al. Science 246:1275-1281 (1989); Winter and Harris, Immunol. Today 14:243-246 (1993); Ward et al., Nature 341:544-546 (1989); Hilyard et al., Protein Engineering: A practical approach (IRL Press 1992); and Borrabeck, Antibody Engineering, 2d ed. (Oxford University Press 1995).

The following example is intended to illustrate but not limit the present invention.

EXAMPLE

This example shows the isolation and characterization of nucleic acid molecules encoding six novel polypeptides having 2,5-DKG permease activity.

Identification of yiaX2 as a 2,5-DKG permease

WO 002170 describes the identification and sequencing of an operon from *Klebsiella oxytoca*, designated the *yia* operon, which contains eight putative open reading frames. Because disruption of this operon abolished the ability of *K. oxytoca* to utilize ascorbic acid as the sole carbon source, the *yia* operon was predicted to be involved in the catabolism of ascorbic acid. The functions of the polypeptides encoded by the individual open reading frames in the *yia* operon were not described in WO 002170.

It was determined that *K. oxytoca* was able to grow on 2,5-DKG as a sole carbon source and, therefore, it was concluded that *K. oxytoca* expressed a 2,5-DKG permease. It was predicted that such a permease would share structural properties with known bacterial transporter proteins, such as multiple transmembrane segments. One of the uncharacterized open reading frames in the *yia* operon, designated *yiaX2*, encoded a transmembrane polypeptide with about 33% identity to a known tartrate transporter, and was thus considered a candidate 2,5-DKG permease.

In order to determine whether *yiaX2* encoded a 2,5-DKG permease, this gene was deleted from the chromosome of the *K. oxytoca* strain designated M5a1. M5a1 has also been

described in the literature as *K. pneumonia* (see, for example, Streicher et al., Proc. Natl. Acad. Sci. 68:1174-1177 (1971)). The *yiaX2* deletion mutant was constructed by joining sequences immediately upstream and downstream of the *yiaX2* gene in a three-way ligation with the pMAK705 integration vector (described in Hamilton et al., J. Bacteriol. 171:4617-4622 (1989)). A fragment of about 1 kb in the *orf1* gene was amplified using oligonucleotides 5'-ACCCAAGCTTCACCAAAAGAGTGAAGAGGAAG-3' (SEQ ID NO:17) and 5'-CGTATCTAGAAAAATATTCTGGTGATGAAGGTGA-3 (SEQ ID NO:18), and digested with HindIII and XbaI. A fragment of a similar size in the *lyxK* gene was amplified with oligonucleotides 5'-AGACTCTAGATCCACATAAACGCACTGCGTAAAC-3' (SEQ ID NO:19) and 5'-GAGGGGATCCTGGCTTCGTGAACGATATACTGG-3' (SEQ ID NO:20), and digested with XbaI and BamHI. The two resulting fragments were ligated together between the HindIII and BamHI sites of the vector pMAK705. The resulting plasmid was transformed into *K. oxytoca* strain M5a1, and candidates in which the deletion construct had integrated by double crossover were obtained as described in Hamilton et al., supra (1989). The designation of the resulting *K. oxytoca* Δ*yiaX2* strain is MGK002. The *yiaX2*-deficient phenotype was verified by by PCR analysis.

As described below, the *K. oxytoca* Δ*yiaX2* [*tkr idnO*] strain was determined to grow very inefficiently on 2,5-DKG as the sole carbon source, and not to grow on 2-KLG. Confirmation that *yiaX2* encoded a polypeptide having 2,5-DKG and 2-KLG permease activities was obtained by determining that adding back the gene restored the ability of the *K. oxytoca* Δ*yiaX2* [*tkr idnO*] to grow well on either 2,5-DKG or 2-KLG (see below).

Construction of K. oxytoca ΔyiaX2 [tkr idnO]

In order to identify additional 2,5-DKG permeases, and preferably permeases selective for 2,5-DKG, a metabolic selection strategy was utilized. As described in WO 00/22170, metabolic selection is advantageous in allowing rapid identification of functional genes from uncharacterized and even unculturable microorganisms, without any prior sequence information.

A tester strain for the metabolic selection of nucleic acid molecules encoding 2,5-DKG permeases was prepared by engineering *K. oxytoca ΔyiaX2* to express enzymes involved in the catabolism of 2,5-DKG to gluconic acid, which can be converted to carbon and energy. Enzymes capable of catabolizing 2,5-DKG to gluconic acid are encoded by the *tkr* and *idnO* genes of the *tkr idnD idnO* operon designated SEQ ID NO:13.

The *tkr idnD idnO* operon (SEQ ID NO:13) was subcloned into the high copy number vector pUC19 and the resulting clone, designated pDF33, was transformed into *K. oxytoca ΔyiaX2*. The resultingtester strain (designated MGK002[pDF33] or *K. oxytoca ΔyiaX2* [*tkr idnO*]) thus expresses all polypeptides required for the utilization of 2,5-DKG as a sole carbon source, but is deficient in 2,5-DKG permease activity to transport extracellular 2,5-DKG into the cell. Therefore, a nucleic acid molecule that encodes a 2,5-DKG permease, upon expression in the tester strain, should confer the ability of the tester strain to grow on

2,5-DKG. The metabolic selection strategy is shown schematically in Figure 3.

To validate the proposed metabolic selection strategy, as a positive control the *yiaX2* gene was reintroduced into the tester strain to confirm that it conferred the ability to grow on 2,5-DKG and 2-KLG. The yiaX2 open reading frame (nucleotides 3777 to 5278 of SEQ ID NO:19 of WO 00/22170) was PCR-amplified using olignucleotides 5'-AATAGGATCCTTCATCACCAGAATATTTTTA-3' (SEQ ID NO:21) and 5'-CATAGGTACCGGCTTTCAGATAGGTGCC-3' (SEQ ID NO:22) digested with BamH1 and Kpn1 and ligated into pCL1920 (Lerner et al., <u>Nucl. Acids. Res</u>. 18:4631 (1990); and see description below) previously digested with the same restriction enzymes. *K. oxytoca* Δ*yiaX2* [*tkr idnO*], transformed with the resulting construct, was able to grow overnight at 30°C on M9 minimal agar medium supplemented with either 2-KLG or 2,5-DKG (0.25%) and 0.1 mM IPTG. Therefore, *K. oxytoca* Δ*yiaX2* [*tkr idnO*] was confirmed to be an appropriate tester strain to identify additional novel 2,5-DKG permeases, and to determine their selectivity.

<u>Construction of bacterial genomic libraries</u>

The cloning vector used for constructing the above positive control and for preparing bacterial genomic libraries is plasmid pCL1920 (Lerner et al., <u>supra</u>, 1990), a low-copy number expression vector which carries a spectinomycin/streptomycin resistance determinant. Expression is driven by the *lac*PO promoter/operator region which is repressed by the *lacI*^{q} gene product when provided

by the host, and induced in the presence of 0.01 to 1mM IPTG.

Genomic DNA from the following species and isolates was prepared according to the method outlined below: *Pantoea citrea* (ATCC 39140), *Klebsiella oxytoca* MGK002 (Δ*yiaX2*), *Pseudomonas aeruginosa*, and a mixture of 25 environmental isolates, obtained from 18 different soil and water samples, and able to grow on 2,5-DKG as the sole carbon source. *Klebsiella oxytoca* MGK002 (Δ*yiaX2*) was among the bacteria chosen because there was a slight amount of background growth observable in the tester strain on 2,5-DKG as the sole carbon source, and some 2,5-DKG permease activity in an uptake assay. However, the tester strain did not grow on 2-KLG, and exhibited no detectable 2-KLG uptake, suggesting the presence of a second 2,5-DKG permease with selectivity for 2,5-DKG in *K. oxytoca*.

Five milliliters of an overnight culture in LB (30°C) were centrifuged for 5 min at 6,000 rpm. Pellets were washed with 1.5 ml Tris 10mM, EDTA 1mM pH 8.0 (TE), centrifuged again and resuspended in 0.4 ml TE. Lysozyme (5mg/ml) and RNase (100pg/ml) were added and cells were incubated for 10 min at 37°C. Sodium dodecylsulfate (SDS) was added to a final concentration of 1% and the tubes were gently shaken until lysis was complete. One hundred microliters of a 5N $NaClO_4$ stock solution were added to the lysate. The mixture was extracted once with one volume of phenol:chloroform (1:1) and once with one volume of chloroform. Chromosomal DNA was precipitated by adding 2 ml of cold (-20°C) ethanol and gently coiling the precipitate around a curved Pasteur pipette. DNA was dried for 30 min

at room temperature and resuspended in 50 to 100 µl of Tris 10mM, EDTA 1mM, NaCl 50mM pH 8.0 to obtain a DNA concentration of 0.5 to 1 µg/µl. Genomic DNA preparations from each environmental isolate were mixed in equal ratios to prepare a single mixed library.

For each preparation, an aliquot of 10-15 µl of genomic DNA was subjected to Sau3A controlled digestion in order to obtain fragments ranging between 3 to 20kb in size. Half that amount was ligated with the low-copy number expression vector pCL1920, which had previously been digested with BamHI and dephosphorylated. The resulting genomic libraries were transformed into *E.coli* DH10B electrocompetent cells (GIBCO-BRL) and briefly amplified overnight at 30°C on LB-agar supplemented with 100µg/ml spectinomycin. For each library, 30,000 to 120,000 clones were plated out and plasmid DNA was bulk-extracted using standard procedures. Insert size was randomly checked and the amplified libraries were stored in the form of plasmid DNA at -20°C for further use in the tester strain.

Selection, identification and sequencing of permease genes

An aliquot of each genomic library was introduced by electroporation into the *K. oxytoca* Δ*yiaX2* [*tkr idnO*] (MGK002[pDF33]) strain. The amount of DNA used in the transformation was adjusted in order to plate out 5 x 10^5 to 1 x 10^6 clones per library on the selective medium. Each selection round was plated on LB-agar containing 100µg/ml spectinomycin, then replica-plated onto M9-agar plates containing 2.5% 2,5-DKG and 0.1 mM IPTG and adjusted to pH 4.5. The clones that grew on 2,5-DKG were transferred into

K. oxytoca ΔyiaX2 (MGK002) devoid of plasmid pDF33, to verify that the *tkr idnDO* pathway was indeed required for growth of those clones on 2,5-DKG. A brief genetic characterization was performed to eliminate identical clones. Following preliminary 2,5-DKG/ 2-KLG uptake assays, 5 clones were retained for further analysis: 2 originated from the *Pantoea citrea* library, 1 from *K. oxytoca* and 2 from the mixed environmental library.

In all cases, DNA sequencing of the vector inserts revealed the presence of a nucleotide sequence (SEQ ID NOS:1, 3, 5, 7 and 9) encoding a polypeptide (SEQ ID NOS:2, 4, 6, 8 and 10) displaying homology with published transporters and with *yiaX2* (SEQ ID NO:11, and its encoded polypeptide SEQ ID NO:12). Also present on these inserts were other orfs, and in most cases an endogenous promoter.

The insert containing both of the prmA and prmB orfs (SEQ ID NOS:7 and 9) was about 9kb, and also contained an orf homologous to bacterial *idnO*, two orfs encoding transcriptional repressors, an orf of unknown function, and 3 orfs encoding homologs of *E. coli* polypeptides involved in nitrate utilization.

The insert containing the PE1 orf (SEQ ID NO:1) was about 3kb, and also contained a putative dehydro-deoxygluconokinase gene closely related to the *B. subtilis* *kdgK* gene and a homolog of the *E. coli* *ydcG* gene.

The insert containing the PE6 orf (SEQ ID NO:3) was about 6.7kb. The genomic environment of PE6 appeared similar to the *yia* operon of *E. coli* and *K. oxytoca*, as SEQ

ID NO:4 was preceded by a *yiaL* homolog and a *yiaK* homolog was also present on the insert.

The insert containing the PK1 orf (SEQ ID NO:5) was about 5.5kb. In contrast to the other inserts, this insert did not appear to contain an endogenous promoter, indicating that the PK1 orf was apparently transcribed from the vector's promoter. The PK1 orf was directly followed by a *tkr* homolog.

Nucleic acid molecules encoding each 2,5-DKG permease were reintroduced into *K. oxytoca* Δ*yiaX2* [*tkr idnO*] and the resulting strains assayed for growth on 2,5-DKG and 2-KLG, and also assayed for uptake of radiolabeled 2,5-DKG and 2-KLG.

The uptake assays were performed by mixing radioactive 2-KLG or 2,5-DKG with IPTG-induced cells, removing aliquots at regular intervals, and measuring both the decrease in radioactivity in the supernatant and the appearance of radioactivity in the cells over time. The results of the growth assays and 2,5-DKG uptake assay are shown in Table 1, below.

Table 1

Recombinantly expressed 2,5 DKG Permease	**Cell Growth on 2,5-DKG**	**Cell Growth on 2-KLG**	**2,5-DKG Uptake (g/l/h)**
YiaX2	+	++	3.7
PE1	++	++	4.2
PE6	++	++	5.0

prmA	++	++	5.5
prmB	+/-	ND	0.9
prmA and prmB	++	ND	9.9
PK1	++	_	4.2
Control *(K. oxytoca ΔyiaX2/ tkr/idn0)*	bkgd	-	1.0

Nucleic acid molecules encoding the different permeases were also subcloned into a variety of vectors, including the high copy number vector pSE380 (which contains a tac promoter), the medium copy number vector pACYC184 (which is promoterless), or the low copy number vector pCL1920, and introduced into a *Pantoea* strain suitable for bioproduction of 2-KLG from glucose (see U.S. Patent No. 5,032,514). The resulting strains were assessed under biofermentation conditions to determine which combinations of nucleic acid molecules, promoters and vectors are optimal for enhancing 2-KLG production.

All journal article, reference and patent citations provided above, in parentheses or otherwise, whether previously stated or not, are incorporated herein by reference in their entirety.

Although the invention has been described with reference to the examples provided above, it should be understood that various modifications can be made without departing from the spirit of the invention.